

Query Rewriting for Retrieval-Augmented Large Language Models

Year: 2023 | Citations: 176 | Authors: Xinbei Ma, Yeyun Gong, Pengcheng He

Abstract

Large Language Models (LLMs) play powerful, black-box readers in the retrieve-then-read pipeline, making remarkable progress in knowledge-intensive tasks. This work introduces a new framework, Rewrite-Retrieve-Read instead of the previous retrieve-then-read for the retrieval-augmented LLMs from the perspective of the query rewriting. Unlike prior studies focusing on adapting either the retriever or the reader, our approach pays attention to the adaptation of the search query itself, for there is inevitably a gap between the input text and the needed knowledge in retrieval. We first prompt an LLM to generate the query, then use a web search engine to retrieve contexts. Furthermore, to better align the query to the frozen modules, we propose a trainable scheme for our pipeline. A small language model is adopted as a trainable rewriter to cater to the black-box LLM reader. The rewriter is trained using the feedback of the LLM reader by reinforcement learning. Evaluation is conducted on downstream tasks, open-domain QA and multiple-choice QA. Experiments results show consistent performance improvement, indicating that our framework is proven effective and scalable, and brings a new framework for retrieval-augmented LLM.